NTP
National Toxicology Program

Levels of Evidence Criteria for NTP Reproductive Toxicology Studies

Edward Carney, Ph.D.
Chair, Reproductive and Developmental
Toxicity Criteria Working Group

NTP Board of Scientific Counselors Meeting

November 20-21, 2008

Charge to the Working Group

- *To evaluate the suitability and utility of proposed criteria for describing the results from individual NTP reproductive toxicology studies to indicate the strength of the evidence for their conclusions.*

- *To evaluate the suitability and utility of proposed criteria for describing the results from individual NTP developmental toxicology studies to indicate the strength of the evidence for their conclusions.*

The Process

- NTP supplied a strawman of the “levels of evidence” criteria to the Board work group (WG) together with an outline of the types of studies conducted by the Program.
- WG undertook an exercise (individually) in applying the criteria to some (~15) study examples selected to explore the boundaries between levels.
- WG reviewed the exercise as a group.
- Made adjustments to the strawman and provided edits on other “key issues” to be used in the application of the criteria.
- Prepared a Work group report.
- Present to the Board in November for action.

Reproductive and Developmental Toxicity Criteria Work group

- Edward Carney (Chair, BSC member)
  - Dow Chemical
- Tracie Bunton (BSC member)
  - EICARTE LLC
- Kenneth Portier (BSC member)
  - American Cancer Society
- Kim Boekelheide (rapporteur reprotox)
  - Brown University
- Robert Chapin
  - Pfizer
- George Daston (rapporteur dev tox)
  - Proctor & Gamble

*Technical Advisors

- James Donald
  - OEHHA
- Earl Gray
  - USEPA
- Barry McIntyre
  - Schering Plough
- Rochelle Tyl
  - RTI International
- Barry Delclos*
  - NCTR, FDA
- Mark Cesta*
  - NTP
- Paul Foster*
  - NTP

Introductory Comments

- The NTP describes the results of individual studies of chemical agents, and notes the strength of the evidence for conclusions regarding each study.
- Negative results, in which the study animals do not exhibit evidence of reproductive toxicity, do not necessarily imply that a chemical is not a reproductive toxicant, but only that the chemical is not a reproductive toxicant under these specific conditions.
- Positive results demonstrating that a chemical causes reproductive toxicity in laboratory animals under the conditions of the study are assumed to be relevant to humans, unless data are available which demonstrate otherwise. In addition, such positive effects should be assumed to be primary effects, unless there is clear evidence that they are secondary consequences of excessive toxicity to non-reproductive organ systems.
- Given that developmental events are intertwined in the reproductive process, effects on developmental toxicity may be detected in reproductive studies. Evaluation of such developmental effects should be based on the NTP Criteria for Levels of Evidence for Developmental Toxicity.

Introductory Comments -2

- It is critical to recognize that the "levels of evidence" statements only describe reproductive **hazard**. The determination of **risk** to humans requires exposure data that are not considered in these summary statements. This is important when communicating study results to the general public.

- Five categories of evidence of reproductive toxicity are used in the NTP Technical Report series to summarize the strength of the evidence observed in each experiment: two categories for positive results **(clear evidence and some evidence);** one category for uncertain findings **(equivocal evidence);** one category for no observable effects **(no evidence);** and one category for experiments that cannot be evaluated because of major design or performance flaws **(inadequate study).**

- In addition, the study's lowest observed adverse effect level is reported for positive results, and the highest dose level tested is reported for the **no evidence** category.

- Application of these criteria requires professional judgment by individuals with ample experience with and understanding of the animal models and study designs employed. For each study, if warranted, these conclusion statements should be made separately for males and females. These categories refer to the weight of evidence of the experimental results and **not** to potency or mechanism.

Levels of Evidence for Reproductive Toxicity - 1

- **Clear Evidence of Reproductive Toxicity**
  - Demonstrated by a dose-related[1] effect on fertility or fecundity, or by changes in multiple interrelated reproductive parameters of sufficient magnitude that by weight of evidence implies a compromise in reproductive function. A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for reproductive toxicity" should accompany the evidence statement.

  - [1]The term "dose-related" describes any dose relationship, recognizing that the treatment-related responses for some endpoints may be non-monotonic due to saturation of exposure or effect, overlapping dose-response behaviors, change in manifestation of the effect at different dose levels, or other phenomena.

Levels of Evidence for Reproductive Toxicity - 2

Some Evidence of Reproductive Toxicity

- Demonstrated by deficits in reproductive parameters, the net impact of which is judged by weight of evidence to have potential to compromise reproductive function. Relative to clear evidence of reproductive toxicity, such effects would be characterized by greater uncertainties or weaker relationships with regard to dose, severity, magnitude, incidence, persistence and/or decreased concordance among affected endpoints.
- A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for reproductive toxicity" should accompany the evidence statement, except in those instances in which the "some" classification has been based on uncertainties about the dose relationship that precludes confident determination of the LOAEL.

Levels of Evidence for Reproductive Toxicity - 3

- **Equivocal Evidence of Reproductive Toxicity**
  - Demonstrated by marginal or discordant deficits in reproductive parameters that may or may not be related to the test article.
- **No Evidence of Reproductive Toxicity**
  - Demonstrated by data from a well conducted, adequate study that are interpreted as showing no biologically relevant evidence of chemically-related deficits in reproductive parameters. A statement to the effect of "This study had no observable adverse reproductive toxicity at the highest dose tested (XXXX mg/kg/d)" should accompany the evidence statement.
- **Inadequate Study of Reproductive Toxicity**
  - Demonstrated by a study that, because of major design or performance flaws, cannot be used to determine the presence of reproductive toxicity.

Key points to consider with the Levels of Evidence criteria

- When a conclusion statement for a particular experiment is selected, consideration must be given to key factors that would extend the boundary of an individual category of evidence. Such consideration should allow for incorporation of scientific experience and current understanding of reproductive toxicity studies in laboratory animals,
  - interrelationships between end points,
  - impact of the change on reproductive function,
  - relative sensitivity of end points, normal background incidence, and specificity of the effect.
- For those evaluations that may be on the borderline between two adjacent levels, some factors to consider in selecting the level of evidence of reproductive toxicity are given below:

- Increases in severity and/or prevalence (more individuals and/or more litters) as a function of dose generally strengthen the level of evidence, keeping in mind that the specific manifestation may change with increasing dose. For example, histological changes at a lower dose level may reflect reductions in fertility at higher dose levels.

Other Key Points -2

- In general, the more animals affected, the stronger the evidence; however, effects on a small number of animals across multiple related endpoints should not be discounted, even in the absence of statistical significance for the individual end point(s).

- Malformations with low incidence should be interpreted in the context of historical controls and may be biologically important.

- Consistency of effects across generations strengthens the level of evidence.

  - Special care should be taken for decrements in reproductive parameters noted in the F_1 generation that were not seen in the F_0 generation, which may suggest developmental as well as reproductive toxicity.

  - Alternatively, if effects are observed in the F_1 generation but not in the F_2 generation (or the effects occur at a lesser frequency in the F_2 generation), this may be due to the nature of the effect resulting in selection (i.e., if the effect is incompatible with successful reproduction, then the affected individuals will not produce offspring).

Other Key Points -3

- Transient changes (e.g., pup weight decrements) by themselves may be weaker indicators of effect than persistent changes.
- Single endpoint changes by themselves may be weaker indicators of effect than concordant effects on multiple interrelated endpoints. -
- Insights from supportive studies (e.g., toxicokinetics, ADME, computational models, structure-activity relationships) and reproductive findings from other *in vivo* animal studies (NTP or otherwise) should be drawn upon when interpreting the biological plausibility of a change.
- New technical approaches and highly sensitive techniques need to have been appropriately characterized to build confidence in their utility, and their usefulness as indicators of effect is increased if they have been anchored to changes in traditional endpoints.
- Clear changes in multiple reproductive tract endpoints without functional changes are sufficient for clear evidence of reproductive toxicity

QUESTIONS?